

#6

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Y. Kato et al.

: Art Unit:

Serial No.: To Be Assigned

: Examiner:

Filed: Herewith

FOR: APPARATUS AND METHOD FOR

PROVIDING INFORMATION BY SPEECH

JC971 U.S. PTO  
09/871283  
05/31/01

## CLAIM TO RIGHT OF PRIORITY

Assistant Commissioner for Patents

Washington, D.C. 20231

S I R :

Pursuant to 35 U.S.C. 119, Applicant's claim to the benefit of filing of prior Japanese Patent Applications No. 2000-163807, filed May 31, 2000 and No. 2000-300061, filed September 29, 2000, as stated in the inventor's Declaration, are hereby confirmed.

A certified copy of each of the above-referenced applications is enclosed.

Respectfully submitted,

  
Allan Ratner, Reg. No. 19,717  
Attorney for Applicants

AR/ap

Enclosures: Certified Copies of Japanese Applications

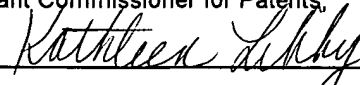
Dated: May 31, 2001

Suite 301  
One Westlakes, Berwyn  
P.O. Box 980  
Valley Forge, PA 19482-0980  
(610) 407-0700

The Assistant Commissioner for Patents is  
hereby authorized to charge payment to  
Deposit Account No. 18-0350 of any fees  
associated with this communication.

**EXPRESS MAIL Mailing Label Number: EL741593262US****Date of Deposit: May 31, 2001**

I hereby certify that this paper and fee are being deposited, under 37 C.F.R. § 1.10 and with sufficient postage, using the "Express Mail Post Office to Addressee" service of the United States Postal Service on the date indicated above and that the deposit is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.



Kathleen Libby

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

JC971 U.S. PTC  
09/871283  
05/31/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されて  
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed  
with this Office

出 願 年 月 日

Date of Application:

2000年 5月31日

出 願 番 号

Application Number:

特願2000-163807

出 願 人

Applicant(s):

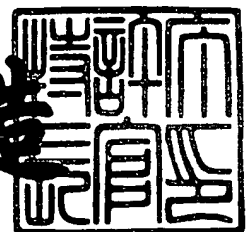
松下電器産業株式会社

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2001年 4月27日

特許庁長官  
Commissioner,  
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3036134

【書類名】 特許願

【整理番号】 2033820019

【提出日】 平成12年 5月31日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 3/00  
G10L 5/02  
G10L 5/04

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 加藤 弓子

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 脇田 由実

【特許出願人】

【識別番号】 000005821

【氏名又は名称】 松下電器産業株式会社

【代理人】

【識別番号】 100092794

【弁理士】

【氏名又は名称】 松田 正道

【電話番号】 06-6397-2840

【手数料の表示】

【予納台帳番号】 009896

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

特 2 0 0 0 - 1 6 3 8 0 7

【物件名】            要約書    1

【包括委任状番号】    9006027

【プルーフの要否】    要

【書類名】 明細書

【発明の名称】 音声による情報提供装置、音声による情報提供方法、媒体及び情報集合体

【特許請求の範囲】

【請求項 1】 単一あるいは複数の入力信号を入力する入力部と、  
前記入力された入力信号の意味を抽出する意味抽出部と、  
前記抽出された意味に対応する標準文型を選択する標準文型検索部と、  
前記抽出された意味に基づきまたは基づかず、前記選択された標準文型の非定型部に当てはまる言葉を生成する非定型部生成部と、  
前記選択された標準文型の定型部の音声情報と、前記生成された言葉の音声情報とを接続し、  
前記接続された音声情報から音声波形を生成する音声合成部と、  
前記生成された音声波形を出力することを特徴とする音声による情報提供装置

【請求項 2】 標準文型とその標準文型の定型部に対応する音声情報とを格納した標準文型データベースを備え、

前記対応する標準文型を選択するとは、前記標準文型検索部が、前記標準文型データベースから前記対応する標準文型の定型部の音声情報を選択することであることを特徴とする請求項 1 記載の音声による情報提供装置。

【請求項 3】 韻律情報を韻律生成単位で格納した非定型部韻律データベースを備え、

前記音声合成部は、前記非定型部韻律データベースを検索して、前記生成された言葉に対応する韻律情報を選択し、その韻律情報に従って前記生成された言葉の前記音声情報を生成することを特徴とする請求項 1 記載の音声による情報提供装置。

【請求項 4】 前記抽出された意味に対応するとは、前記抽出された意味と同等の意味を持つことであることを特徴とする請求項 2 記載の音声による情報提供装置。

【請求項 5】 前記抽出された意味に対応するとは、前記抽出された意味に対

する応答であることであることを特徴とする請求項 2 記載の音声による情報提供装置。

【請求項 6】 前記標準文型データベースが、前記対応する標準文型の定型部に対応する動作及び／または画像データをも格納していることを特徴とする請求項 5 記載の音声による情報提供装置。

【請求項 7】 前記標準文型データベースが、前記対応する標準文型の定型部に対応する動作をも格納している場合であって、

前記標準文型検索部は、前記抽出された意味に対する応答である標準文型の定型部に対応する動作をも選択し、

前記生成された音声波形の出力に対応して前記選択された動作が行われることを特徴とする請求項 6 記載の音声による情報提供装置。

【請求項 8】 前記標準応答文型データベースが、前記対応する標準文型に対応する画像データをも格納している場合であって、

前記標準文型検索部は、前記抽出された意味に対する応答である標準文型の定型部に対応する画像をも選択し、

前記生成された音声波形の出力に対応して前記選択された画像が表示されることを特徴とする請求項 6 記載の音声による情報提供装置。

【請求項 9】 前記入力信号は、テキスト、音声、音、画像、振動、加速度、温度、張力の少なくとも一つ以上であることを特徴とする請求項 1 記載の音声による情報提供装置。

【請求項 10】 前記入力信号が少なくとも音声である場合であって、

前記入力された音声を認識する音声認識部を備え、

前記意味抽出部は、前記認識された結果に基づいて、前記入力信号の意味を抽出することを特徴とする請求項 9 記載の音声による情報提供装置。

【請求項 11】 前記入力信号が少なくとも音である場合であって、

前記入力された音の音源を認識する音源認識部を備え、

前記意味抽出部は、前記認識された結果に基づいて、前記入力信号の意味を抽出することを特徴とする請求項 9 記載の音声による情報提供装置。

【請求項 12】 前記入力信号が少なくとも画像である場合であって、

前記入力された画像を認識するに画像認識部を備え、

前記意味抽出部は、前記認識された結果に基づいて、前記入力信号の意味を抽出することを特徴とする請求項 9 記載の音声による情報提供装置。

【請求項 1 3】 前記標準文型データベースの標準文型の定型部に対応する音声情報は、自然発声を録音した音声波形であることを特徴とする請求項 2 記載の音声による情報提供装置。

【請求項 1 4】 前記標準文型データベースの標準文型の定型部に対応する音声情報は、自然発声音声より抽出した韻律情報であることを特徴とする請求項 2 記載の音声による情報提供装置。

【請求項 1 5】 前記韻律情報は、音声の基本周波数パターン、強度パターン、音韻持続時間長パターンのうち少なくとも1つを含むことを特徴とする請求項 3 または 1 4 に記載の音声による情報提供装置。

【請求項 1 6】 前記非定型部韻律データベースは、音韻列、モーラ数、音節数、アクセント、文中での位置、直前あるいは直後のポーズの有無および時間長、直前あるいは直後のアクセント句のアクセント型、プロミネンス、品詞列、文節属性および係り受け関係のうち少なくとも1つの条件とそれに対応する韻律情報を格納していることを特徴とする請求項 3 記載の音声による情報提供装置。

【請求項 1 7】 前記韻律生成単位とは、アクセント句、フレーズ、単語、文節あるいは節のいずれかであることを特徴とする請求項 3 記載の音声による情報提供装置。

【請求項 1 8】 単一あるいは複数の入力信号を入力し、  
前記入力信号の意味を抽出し、  
前記抽出された意味に対応する標準文型を選択し、  
前記抽出された意味に基づきまたは基づかず、前記選択された標準文型の非定型部に当てはまる言葉を生成し、

前記選択された標準文型の定型部の音声情報と、前記生成された言葉の音声情報とを接続し、

前記接続された音声情報から音声波形を生成し、

前記生成された音声波形を出力することを特徴とする音声による情報提供方法

【請求項 1 9】 前記抽出された意味に対応するとは、前記抽出された意味と同等の意味を持つことであることを特徴とする請求項 1 8 記載の音声による情報提供装置。

【請求項 2 0】 前記抽出された意味に対応するとは、前記抽出された意味に対する応答であることであることを特徴とする請求項 1 8 記載の音声による情報提供装置。

【請求項 2 1】 請求項 1 ～ 1 7 のいずれかに記載の音声による情報提供装置の各部の全部または一部の全部または一部の機能をコンピュータにより実行させるためのプログラム及び／またはデータを担持した媒体であって、コンピュータにより処理可能なことを特徴とする媒体。

【請求項 2 2】 請求項 1 ～ 1 7 のいずれかに記載の音声による情報提供装置または対話装置の各部の全部または一部の全部または一部の機能をコンピュータにより実行させるためのプログラム及び／またはデータであることを特徴とする情報集合体。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、入力された任意のテキスト、音声、画像等の入力信号を解釈して音声に変換して出力する音声による情報提供装置、音声による情報提供方法および入力された任意のテキスト、音声、画像等の入力信号を解釈してその入力信号に対応する応答を出力する対話システム、対話制御方法、および媒体、情報集合体に関するものである。

【0 0 0 2】

【従来の技術】

従来の音声による情報提供は、交通情報や天気概況の音声サービスのような定型文を聞き取りやすく自然な音声で読み上げるために、特開平 8 - 6 3 1 8 7 の音声合成装置のように合成すべき一群のメッセージのすべてに共通する固定情報である定型部とメッセージ毎に異なる非定型部に分け、定型部にはデータベース



内に格納された音韻持続時間長と基本周波数パターンを適用し、非定型部には非定型部の文中位置毎に音節数とアクセント型で分類、格納された音韻持続時間長と基本周波数パターンを適用し、両者の基本周波数を接続し、それらの情報に基づいて音声波形を生成していた。

【 0 0 0 3 】

【発明が解決しようとする課題】

しかしながら、このような構成では提供しようとする情報は必ず定型文の形式のテキストにしなければならず、定型部データベースに格納されている表現以外は合成することは不可能である。

【 0 0 0 4 】

つまり、定型部データベースに格納されたテキスト表現以外の入力を受け付けず、使用者は入力テキスト作成時に定型部データベース内の表現を参照する必要があり、自由な入力ができず不便である。

【 0 0 0 5 】

すなわち、従来の音声による情報提供方法では、自由な入力を行うことが出来ないという課題がある。

【 0 0 0 6 】

本発明は、上記課題を考慮し、任意のテキストの入力を受け付けるのみでなく、音声、画像、音響等自由な入力信号を受け付けることが出来、意味を解釈して言語表現としての標準文型へ変換し、自然な音声で情報を提供することが出来る音声による情報提供システム、情報提供方法、対話システム、対話制御方法、媒体及び情報集合体を提供することを目的とするものである。

【 0 0 0 7 】

【課題を解決するための手段】

上述した課題を解決するために、第 1 の本発明（請求項 1 に対応）は、  
単一あるいは複数の入力信号を入力する入力部と、  
前記入力された入力信号の意味を抽出する意味抽出部と、  
前記抽出された意味に対応する標準文型を選択する標準文型検索部と、  
前記抽出された意味に基づきまたは基づかず、前記選択された標準文型の非定

型部に当てはまる言葉を生成する非定型部生成部と、

前記選択された標準文型の定型部の音声情報と、前記生成された言葉の音声情報とを接続し、

前記接続された音声情報から音声波形を生成する音声合成部と、

前記生成された音声波形を出力することを特徴とする音声による情報提供装置である。

【 0 0 0 8 】

また、第 2 の本発明（請求項 2 に対応）は、標準文型とその標準文型の定型部に対応する音声情報とを格納した標準文型データベースを備え、

前記対応する標準文型を選択するとは、前記標準文型検索部が、前記標準文型データベースから前記対応する標準文型の定型部の音声情報を選択することであることを特徴とする第 1 の本発明に記載の音声による情報提供装置である。

【 0 0 0 9 】

また、第 3 の本発明（請求項 3 に対応）は、韻律情報を韻律生成単位で格納した非定型部韻律データベースを備え、

前記音声合成部は、前記非定型部韻律データベースを検索して、前記生成された言葉に対応する韻律情報を選択し、その韻律情報に従って前記生成された言葉の前記音声情報を生成することを特徴とする第 1 の本発明に記載の音声による情報提供装置である。

【 0 0 1 0 】

また、第 4 の本発明（請求項 4 に対応）は、前記抽出された意味に対応するとは、前記抽出された意味と同等の意味を持つことであることを特徴とする第 2 の本発明に記載の音声による情報提供装置である。

【 0 0 1 1 】

また、第 5 の本発明（請求項 5 に対応）は、前記抽出された意味に対応するとは、前記抽出された意味に対する応答であることであることを特徴とする第 2 の本発明に記載の音声による情報提供装置である。

【 0 0 1 2 】

また、第 6 の本発明（請求項 6 に対応）は、前記標準文型データベースが、前

記対応する標準文型の定型部に対応する動作及び／または画像データをも格納していることを特徴とする第 5 の本発明に記載の音声による情報提供装置である。

【 0 0 1 3 】

また、第 7 の本発明（請求項 7 に対応）は、前記標準文型データベースが、前記対応する標準文型の定型部に対応する動作をも格納している場合であって、

前記標準文型検索部は、前記抽出された意味に対する応答である標準文型の定型部に対応する動作をも選択し、

前記生成された音声波形の出力に対応して前記選択された動作が行われることを特徴とする第 6 の本発明に記載の音声による情報提供装置である。

【 0 0 1 4 】

また、第 8 の本発明（請求項 8 に対応）は、前記標準応答文型データベースが、前記対応する標準文型に対応する画像データをも格納している場合であって、

前記標準文型検索部は、前記抽出された意味に対する応答である標準文型の定型部に対応する画像をも選択し、

前記生成された音声波形の出力に対応して前記選択された画像が表示されることを特徴とする第 6 の本発明に記載の音声による情報提供装置である。

【 0 0 1 5 】

また、第 9 の本発明（請求項 9 に対応）は、前記入力信号は、テキスト、音声、音、画像、振動、加速度、温度、張力の少なくとも一つ以上であることを特徴とする第 1 の本発明に記載の音声による情報提供装置である。

【 0 0 1 6 】

また、第 1 0 の本発明（請求項 1 0 に対応）は、前記入力信号が少なくとも音声である場合であって、

前記入力された音声を認識する音声認識部を備え、

前記意味抽出部は、前記認識された結果に基づいて、前記入力信号の意味を抽出することを特徴とする第 9 の本発明に記載の音声による情報提供装置である。

【 0 0 1 7 】

また、第 1 1 の本発明（請求項 1 1 に対応）は、前記入力信号が少なくとも音である場合であって、

前記入力された音の音源を認識する音源認識部を備え、

前記意味抽出部は、前記認識された結果に基づいて、前記入力信号の意味を抽出することを特徴とする第 9 の本発明に記載の音声による情報提供装置である。

【 0 0 1 8 】

また、第 1 2 の本発明（請求項 1 2 に対応）は、前記入力信号が少なくとも画像である場合であって、

前記入力された画像を認識するに画像認識部を備え、

前記意味抽出部は、前記認識された結果に基づいて、前記入力信号の意味を抽出することを特徴とする第 9 の本発明に記載の音声による情報提供装置である。

【 0 0 1 9 】

また、第 1 3 の本発明（請求項 1 3 に対応）は、前記標準文型データベースの標準文型の定型部に対応する音声情報は、自然発声を録音した音声波形であることを特徴とする第 2 の本発明に記載の音声による情報提供装置である。

【 0 0 2 0 】

また、第 1 4 の本発明（請求項 1 4 に対応）は、前記標準文型データベースの標準文型の定型部に対応する音声情報は、自然発声音声より抽出した韻律情報であることを特徴とする第 2 の本発明に記載の音声による情報提供装置である。

【 0 0 2 1 】

また、第 1 5 の本発明（請求項 1 5 に対応）は、前記韻律情報は、音声の基本周波数パターン、強度パターン、音韻持続時間長パターンのうち少なくとも 1 つを含むことを特徴とする第 3 または 1 4 の本発明に記載の音声による情報提供装置である。

【 0 0 2 2 】

また、第 1 6 の本発明（請求項 1 6 に対応）は、前記非定型部韻律データベースは、音韻列、モーラ数、音節数、アクセント、文中での位置、直前あるいは直後のポーズの有無および時間長、直前あるいは直後のアクセント句のアクセント型、プロミネンス、品詞列、文節属性および係り受け関係のうち少なくとも 1 つの条件とそれに対応する韻律情報を格納していることを特徴とする第 3 の本発明に記載の音声による情報提供装置である。

【 0 0 2 3 】

また、第 1 7 の本発明（請求項 1 7 に対応）は、前記韻律生成単位とは、アクセント句、フレーズ、単語、文節あるいは節のいずれかであることを特徴とする第 3 の本発明に記載の音声による情報提供装置である。

【 0 0 2 4 】

また、第 1 8 の本発明（請求項 1 8 に対応）は、単一あるいは複数の入力信号を入力し、

前記入力信号の意味を抽出し、

前記抽出された意味に対応する標準文型を選択し、

前記抽出された意味に基づきまたは基づかず、前記選択された標準文型の非定型部に当てはまる言葉を生成し、

前記選択された標準文型の定型部の音声情報と、前記生成された言葉の音声情報とを接続し、

前記接続された音声情報から音声波形を生成し、

前記生成された音声波形を出力することを特徴とする音声による情報提供方法である。

【 0 0 2 5 】

また、第 1 9 の本発明（請求項 1 9 に対応）は、前記抽出された意味に対応するとは、前記抽出された意味と同等の意味を持つことであることを特徴とする第 1 8 の本発明に記載の音声による情報提供装置である。

【 0 0 2 6 】

また、第 2 0 の本発明（請求項 2 0 に対応）は、前記抽出された意味に対応するとは、前記抽出された意味に対する応答であることであることを特徴とする第 1 8 の本発明に記載の音声による情報提供装置である。

【 0 0 2 7 】

また、第 2 1 の本発明（請求項 2 1 に対応）は、第 1 ～ 1 7 の本発明のいずれかに記載の音声による情報提供装置の各部の全部または一部の全部または一部の機能をコンピュータにより実行させるためのプログラム及び／またはデータを担持した媒体であって、コンピュータにより処理可能なことを特徴とする媒体であ

る。

【 0 0 2 8 】

また、第 2 2 の本発明（請求項 2 2 に対応）は、第 1 ～ 1 7 の本発明のいずれかに記載の音声による情報提供装置または対話装置の各部の全部または一部の全部または一部の機能をコンピュータにより実行させるためのプログラム及び／またはデータであることを特徴とする情報集合体である。

【 0 0 2 9 】

【発明の実施の形態】

以下に、本発明の実施の形態を図面を参照して説明する。

【 0 0 3 0 】

（実施の形態 1）

図 1 は、本発明の実施の形態 1 における音声による情報提供システムの構成を示す機能ブロック図である。また、図 2 に本発明の実施の形態 1 における音声による情報提供システムの動作を示す流れ図を示す。

【 0 0 3 1 】

図 1 において、110 はテキストを入力する入力部である。120 は表記、読み、品詞、意味タグ等を格納する言語処理用辞書であり、130 は言語処理用辞書 120 を参照して入力テキストの形態素解析および意味タグの付与を行う意味抽出部である。140 は標準文型ごとの意味タグ列、定型部音素列、定型部韻律パターン、非定型部韻律パターンの調整パラメータを格納する標準文型データベースであり、150 は意味タグ列を用いて標準文型データベースを検索する標準文型検索部であり、160 は非定型部に当たる入力の表音記号列を生成する非定型部生成部である。170 は音声合成部であり、180 は音声波形を出力する出力部である。音声合成部 170 は音素列、モーラ数、アクセント、文中での位置、直前直後のポーズの有無および時間長、直前直後のアクセント句のアクセント型と韻律情報を格納した非定型部韻律データベース 171 と、非定型部韻律データベース 171 を参照して非定型部の韻律情報を抽出し、標準文型検索部 150 で抽出された定型部韻律情報と接続する韻律制御部 172 と、波形生成ユニットを格納した音声素変データベース 173 と音声素変データベース 173 に格納された

音声素変を用いて韻律制御部 172 より出力される韻律情報に基づいて音声波形を生成する波形生成部 174 を含む。

#### 【0032】

以上のように構成された音声による情報提供システムの動作を図 2 に従って説明する。

#### 【0033】

入力部 110 は、処理すべきテキストデータを受け付け（ステップ 10）、意味抽出部 130 は、入力されたテキストデータを言語処理用辞書 120 を用いて形態素解析して、読み、品詞等の言語情報を付与し、さらに構文解析を行い、構文単位ごとに意味タグを付与する（ステップ 20）。

#### 【0034】

このようなステップ 20 の動作を図 8 を用いて説明する。入力テキストが入力テキスト 400、すなわち「A 氏いわく、「芸術は爆発だ」」であるとする。このテキストデータを形態素解析すると、形態素解析結果 401 のようになる。そして、各形態素に読み、品詞などの言語情報を付与する。例えば形態素「A」は、読みが「エイ」で品詞が名詞であり、形態素「氏」は、読みが「シ」で品詞は「接尾詞」となる。次に言語情報が付与された形態素解析結果 401 の構文解析を行い、構文単位毎に意味タグを付与すると、意味タグ付与結果 402 のようになる。本実施の形態では、構文単位として文節を用いた。すなわち、「A 氏」には意味タグ「who」が付与され、「いわく、」には「action 言う」が付与される。また、引用文の部分は一つの文節とみなし「「芸術は爆発だ」」には、「what 引用」を付与する。

#### 【0035】

標準文型検索部 150 は、意味抽出部 130 から出力された構文単位ごとの意味タグ列を用いて標準文型データベース 140 を検索し（ステップ 30）、入力テキストを特定の標準文型にマッピングし、マッピングされた標準文型の定型部の音素列と韻律情報を抽出する（ステップ 40）。

#### 【0036】

またステップ 40 でマッピングされた標準文型の非定型部属性とステップ 20

で付与された言語情報を比較し、入力テキストから非定型部にあたる言葉を生成する(ステップ50)。

#### 【0037】

韻律制御部172はステップ50で生成された非定型部の音素列、モーラ数、アクセントと、非定型部の文中での位置、非定型部と定型部の間のポーズの有無および時間長、非定型部に隣接する定型部のアクセント型を用いて非定型部韻律データベース171を検索し(ステップ60)、非定型部の韻律情報をアクセント句単位で抽出する(ステップ70)。

#### 【0038】

さらに韻律制御部172はステップ40でマッピングされた標準文型の非定型部韻律調整パラメータに基づいてステップ60で抽出された非定型部の韻律情報を調整し、ステップ40で抽出された定型部の韻律情報と接続する。調整は例えば図3(A)あるいは図3(B)のように行う(ステップ80)。

#### 【0039】

波形生成部174はステップ40で抽出された定型部の音素列とステップ50で生成された非定型部の音素列およびステップ80で生成された韻律情報に基づき音声素変データベース173に格納された音声素変を利用して音声波形を生成する(ステップ90)。

#### 【0040】

ステップ90で生成された音声波形を出力部180より音声として出力する(ステップ100)。

#### 【0041】

以上のように、本実施の形態の音声による情報提供システムによれば入力テキストの意味を抽出し、その意味が同等である標準文型に変換して音声を合成することにより、任意のテキストに対して、定型文を利用することで自然性の高い合成音声を実現する音声合成部を利用し、自然な音声で正確に情報を提供することができる。

#### 【0042】

さらに、任意のテキストの入力を入力しても自然な音声で正確に情報を提供す



ることが出来る。

#### 【 0 0 4 3 】

##### （実施の形態 2）

図 4 は、本発明の実施の形態 2 における音声による情報提供システムの構成を示す機能ブロック図である。また、図 5 に本発明の実施の形態 1 における音声による情報提供システムの動作を示す流れ図を示す。

#### 【 0 0 4 4 】

図 4 において図 1 と同一な部分については同一符号を付しているので説明を省略し、異なった部分についてのみ説明する。実施の形態 2 の図 4 は図 1 の構成の言語処理用辞書 1 2 0 が英語の言語処理に用いる英語辞書 2 2 0 に置き換わり、意味抽出部 1 3 0 が入力音声を認識し、英語辞書 2 2 0 を参照して意味タグの付与を行う音声認識／意味抽出部 2 3 0 に置き換わり、意味タグと意味タグに対応する日本語を格納した日本語辞書 2 2 5 が付け加わり、非定型部生成部 1 6 0 が日本語辞書 2 2 5 を参照して非定型部の日本語を生成する非定型部日本語生成部 2 6 0 に置き換わった以外はは実施の形態 1 の構成と同一な構成である。

#### 【 0 0 4 5 】

なお、本実施の形態の音声認識／意味抽出部 2 3 0 は本発明の音声認識部の例であり、本実施の形態の音声認識／意味抽出部 2 3 0 は本発明の意味抽出部の例を兼ねており、本実施の形態の非定型部日本語生成部 2 6 0 は本発明の非定型部生成部の例である。

#### 【 0 0 4 6 】

以上のように構成された音声による情報提供システムの動作を図 5 に従って説明する。

#### 【 0 0 4 7 】

入力部 1 1 0 は、処理すべき英語音声波形を受け付け（ステップ 1 1 0）、音声認識／意味抽出部 2 3 0 は、入力された音声を認識し、英語の単語列に変換する（ステップ 1 1 5）。さらに音声認識／意味抽出部 2 3 0 は音声認識結果に対して、形態素解析し、品詞等の言語情報を付与し、さらに構文解析を行い、構文単位ごとに意味タグを付与する（ステップ 1 2 0）。

## 【 0 0 4 8 】

なお、ステップ 1 2 0 は第 1 の実施の形態で図 8 を用いて説明した動作例と同様の動作をする。

## 【 0 0 4 9 】

標準文型検索部 1 5 0 は、音声認識／意味抽出部 2 3 0 から出力された構文単位ごとの意味タグ列を用いて標準文型データベース 1 4 0 を検索し(ステップ 1 3 0)、音声認識結果を特定の日本語標準文型にマッピングし、マッピングされた日本語標準文型の定型部の音素列と韻律情報を抽出する(ステップ 1 4 0)。

## 【 0 0 5 0 】

またステップ 1 4 0 でマッピングされた日本語標準文型の非定型部属性とステップ 1 2 0 で付与された意味タグに対応する日本語を日本語辞書 2 5 5 より抽出し、非定型部にあたる日本語を生成する(ステップ 1 5 0)。

## 【 0 0 5 1 】

以下ステップ 6 0 からステップ 1 0 0 までは実施の形態 1 において図に基づいて説明したのと同様にし、日本語音声を出力する。

## 【 0 0 5 2 】

以上のように、本実施の形態の音声による情報提供システムによれば入力された英語音声の意味を抽出し、その意味が同等である日本語の標準文型に変換して音声を合成することにより、任意のテキストに対して、定型文を利用することで自然性の高い合成音声を実現する音声合成部を利用し、自然な音声で簡便に通訳情報を提供することができる。

## 【 0 0 5 3 】

## (実施の形態 3)

図 6 は、本発明の実施の形態 3 における音声による情報提供システムの構成を示す機能ブロック図である。また、図 7 に本発明の実施の形態 3 における音声による対話システムの動作を示す流れ図を示す。

## 【 0 0 5 4 】

図 6 において図 1 と同一な部分については同一符号を付しているので説明を省略し、異なった部分についてのみ説明する。

## 【0055】

実施の形態3の図6は図1の構成の入力部110が音声を入力する音声入力部311と画像を入力する画像入力部312に置き換わり、言語処理用辞書120が音声および画像の特徴量と意味タグを格納した認識辞書320に置き換わり、意味抽出部130が音声入力部311から入力された音声と画像入力部312から入力された画像を認識辞書320を参照して音声及び画像の認識を行い意味タグを付与する認識／意味抽出部330に置き換わり、標準文型データベース140が標準化された応答音声の定型部情報と応答画像のタグを格納した標準応答データベース340に置き換わり、標準文型検索部150が意味タグ列を用いて、対応する応答を標準応答データベースより検索、抽出する応答表現検索部350に置き換わり、出力部180が音声出力部381と画像出力部382に置き換わり、応答画像を格納した画像データベース371と、応答表現検索部350で標準応答データベース340より抽出された画像タグに基づき画像データを生成する画像生成部370と、音声出力と画像出力のタイミングを調整するタイミング制御部380が付け加わった以外は実施の形態1の構成と同一な構成である。

## 【0056】

なお、本実施の形態の音声入力部311、画像入力部312は本発明の入力部の例であり、本実施の形態の認識／意味抽出部330は本発明の音声認識部の例であり、本実施の形態の認識／意味抽出部330は本発明の意味抽出部の例を兼ねており、本実施の形態の標準応答データベース340、画像データベース371は本発明の標準文型データベースの例である。

## 【0057】

以上のように構成された音声による情報提供システムの動作を図7に従って説明する。

## 【0058】

音声入力部311は処理すべき音声波形を受け付け、画像入力部312は処理すべき音声に同期した画像データを受け付け（ステップ210）、認識／意味抽出部330は、入力された音声を認識し単語列に変換し、入力された画像を認識して画像タグを付与する（ステップ215）。さらに認識／意味抽出部330は

単語列を形態素解析し、品詞等の言語情報を付与し、構文解析を行い、音声の処理結果と画像の処理結果をあわせて構文単位ごとに意味タグを付与する（ステップ 2 2 0）。

【 0 0 5 9 】

なお、ステップ 2 2 0 は、第 1 の実施の形態で図 8 を用いて説明した動作例と同様の動作をする。

【 0 0 6 0 】

応答表現検索部 3 5 0 は、認識／意味抽出部 3 3 0 から出力された構文単位ごとの意味タグ列および画像タグを用いて標準応答データベース 3 4 0 を検索し、音声および画像の認識結果に対する応答をマッピングする（ステップ 2 3 0）。

【 0 0 6 1 】

さらにマッピングされた応答の標準応答文型の定型部の音素列と韻律情報を抽出する（ステップ 2 4 0）。

【 0 0 6 2 】

またマッピングされた応答の画像タグと画像と標準応答文型との同期情報を抽出する（ステップ 3 4 0）。

【 0 0 6 3 】

非定型部生成部 1 7 0 はステップ 2 4 0 でマッピングされた標準応答文型の非定型部属性とステップ 2 2 0 で付与された意味タグに対応する単語あるいはフレーズを認識辞書 3 2 0 より抽出し、非定型部を生成する（ステップ 2 5 0）。

【 0 0 6 4 】

以下ステップ 6 0 からステップ 9 0 までは実施の形態 1 において図 2 に基づいて説明したのと同様にし、音声波形を生成する。

【 0 0 6 5 】

一方、画像生成部 3 7 0 はステップ 2 3 0 でマッピングされた応答の応答画像タグを用いて画像データベース 3 7 1 より応答画像を抽出し（ステップ 3 6 0）、標準応答文型との同期情報に基づいて画像を生成する（ステップ 3 8 0）。

【 0 0 6 6 】

タイミング制御部 3 8 0 はステップ 3 4 0 で抽出された応答画像と標準応答文

型との同期情報に基づき、ステップ90で生成された音声波形とステップ380で生成された画像の同期をとり、音声出力部381および画像出力部382より応答の音声と画像を出力する。

## 【0067】

以上のように、本実施の形態の対話システムによれば入力された音声と画像の意味を抽出し、その意味に対する標準応答文型に基づいて応答音声を合成するとともに応答画像を生成することにより、任意の入力に対して、語順や表現のバリエーションに関わらず効率よく応答文を生成できる上に、定型文を利用することで自然性の高い合成音声を実現する音声合成部を利用し、自然な音声で対話応答を生成することが出来る。

## 【0068】

なお、実施の形態1、2および3において、音声合成は音声素変を接続して行う方法としてが、これ以外の方法でもよい。

## 【0069】

なお、実施の形態1および2の標準文型データベースおよび実施の形態3の標準応答データベースは定型部音素列、定型部韻律パターン、非定型部韻律パターンの調整パラメータを格納するとしたが、定型部音素列、定型部韻律パターンのかわりに、録音音声を格納するとしてもよい。

## 【0070】

なお、実施の形態1および2において標準文型データベースおよび実施の形態3の標準データベースは定型部音素列、定型部韻律パターン、非定型部韻律パターンの調整パラメータを格納するとしたが、定型部音素列、定型部韻律パターンのかわりに、音声合成部170での合成方式にあわせたホルマント情報等パラメータを格納するとしてもよい。

## 【0071】

なお、実施の形態1、2および3において、非定型部韻律データベース171は音素列モーラ数、アクセント、文中での位置、直前直後のポーズの有無および時間長、直前直後のアクセント句のアクセント型と韻律情報を格納するとしたが、これ以外に品詞列、文節属性、係り受け、プロミネンス等を格納するあるいは

上記のうち韻律情報以外に少なくとも1つを格納すればよい。

【0072】

なお、実施の形態1および2において、入力は単一の信号としたが複数の入力信号を受け付けるものとしてもよい。

【0073】

なお、実施の形態3において、入力は複数の信号としたが単一の入力信号を受け付けるものとしてよい。

【0074】

なお、実施の形態1において入力はテキストとしたが、テキスト以外の音声、音、画像、振動、加速度、温度、張力等のいずれかあるいはそれらの組み合わせであってもよい。

【0075】

なお、実施の形態2において入力は音声としたが、テキスト、あるいは音声とテキストの組み合わせであってもよい。

【0076】

なお、実施の形態3において入力は音声と画像としたが、それ以外の音、振動、加速度、温度、張力等のいずれかあるいはその組み合わせであってもよい。

【0077】

なお、実施の形態2において英語を日本語に変換するものとしたがそれ以外の言語であってもよい。

【0078】

なお、実施の形態2において入力音声の言語を単一の言語としたが、自動あるいは使用者の選択により複数言語を切り替えるものとしてもよい。

【0079】

なお、実施の形態2において出力音声の言語を単一の言語としたが、使用者の選択により複数言語を切り替えるものとしてもよい。

【0080】

このように本実施の形態によれば、テキスト、音声、画像その他の自由な入力に対し、入力信号の意味を解釈し標準文型による言語表現に変換することで広範

なメディア、モダリティから音声への変換および言語変換を可能にし、かつ高品質な音声で情報を提供することが出来る。

## 【 0 0 8 1 】

なお、本発明の音声による情報提供装置または対話装置の各部の全部または一部の各部の全部または一部の機能をコンピュータにより実行させるためのプログラム及び／またはデータを担持した媒体であって、コンピュータにより処理可能なことを特徴とする媒体も本発明に属する。

## 【 0 0 8 2 】

さらに、本発明の音声による情報提供装置または対話装置の各部の全部または一部の全部または一部の機能をコンピュータにより実行させるためのプログラム及び／またはデータであることを特徴とする情報集合体も本発明に属する。

## 【 0 0 8 3 】

さらに、本発明のデータとは、データ構造、データフォーマット、データの種類などを含む。また、本発明の媒体とは、ROM等の記録媒体、インターネット等の伝送媒体、光・電波・音波等の伝送媒体を含む。また、本発明の担持した媒体とは、例えば、プログラム及び／またはデータを記録した記録媒体、やプログラム及び／またはデータを伝送する伝送媒体等を含む。また、本発明のコンピュータにより処理可能とは、例えば、ROMなどの記録媒体の場合であれば、コンピュータにより読みとり可能であることであり、伝送媒体の場合であれば、伝送対象となるプログラム及び／またはデータが伝送の結果として、コンピュータにより取り扱えることであることを含む。また、本発明の情報集合体とは、例えば、プログラム及び／またはデータ等のソフトウェアを含むものである。

## 【 0 0 8 4 】

さらに、上記実施の形態の音声による情報提供システムの各部の全部または一部の全部または一部の機能をコンピュータにより実行させるためのプログラム及び／またはデータを記録したプログラム記録媒体は、コンピュータにより読み取り可能であり、読み取られた前記プログラム及び／またはデータが前記コンピュータと協働して前記機能を実行するプログラム記録媒体であっても良い。

## 【 0 0 8 5 】

【発明の効果】

本発明は、任意のテキストの入力を受け付けるのみでなく、音声、画像、音響等自由な入力信号を受け付けることが出来、自然な音声で情報を提供することが出来る音声による情報提供システム、情報提供方法、対話システム、対話制御方法、媒体及び情報集合体を提供することが出来る。

【図面の簡単な説明】

【図 1】

本発明の実施の形態 1 の音声による情報提供システムのブロック図

【図 2】

本発明の実施の形態 1 の動作を示す流れ図

【図 3】

本発明の実施の形態 1 の韻律情報の接続方法の例を示す図

【図 4】

本発明の実施の形態 2 の音声による情報提供システムのブロック図

【図 5】

本発明の実施の形態 2 の動作を示す流れ図

【図 6】

本発明の実施の形態 3 の対話システムのブロック図

【図 7】

本発明の実施の形態 3 の動作を示す流れ図

【図 8】

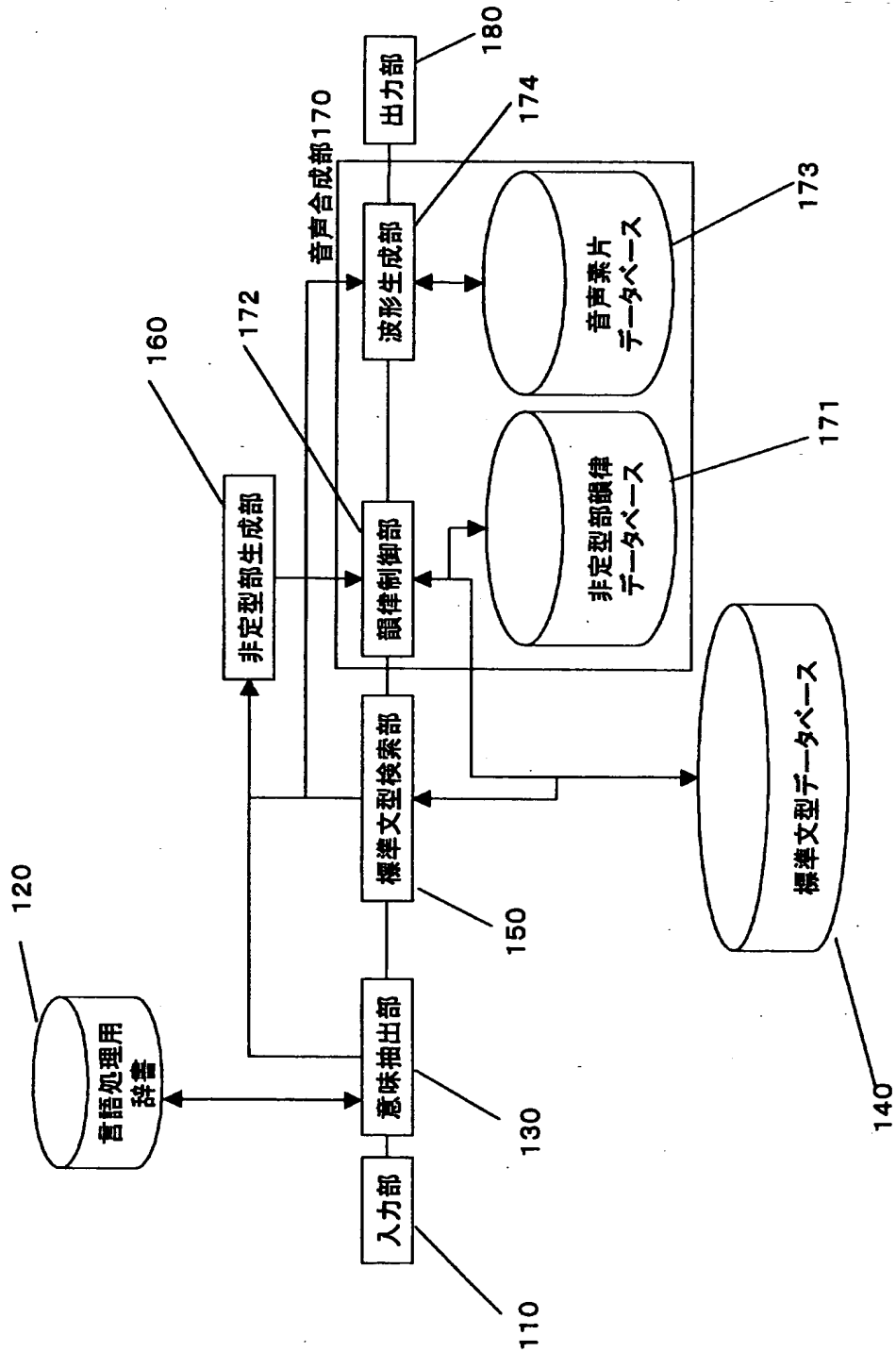
本発明の実施の形態 1 ～ 3 における意味タグを付与する動作を説明する図



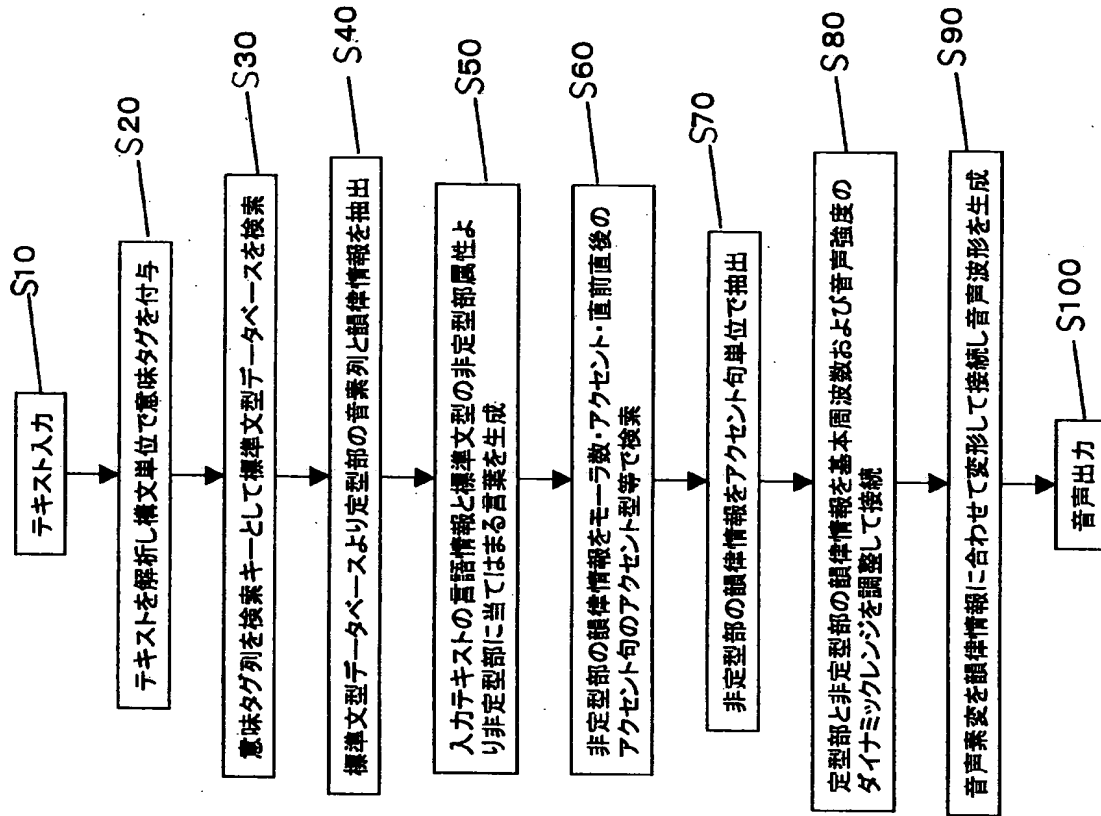
【書類名】

図面

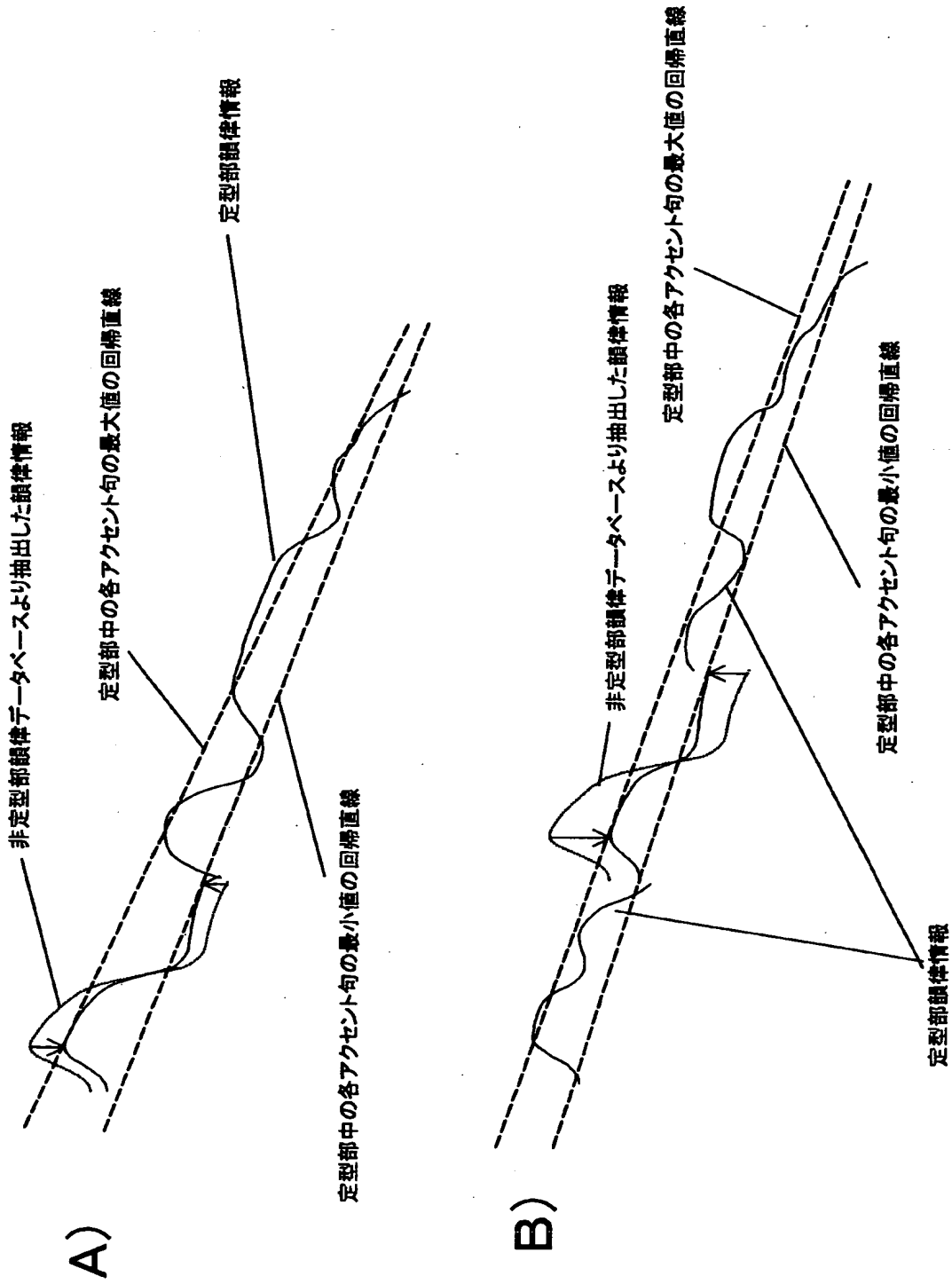
【図 1】



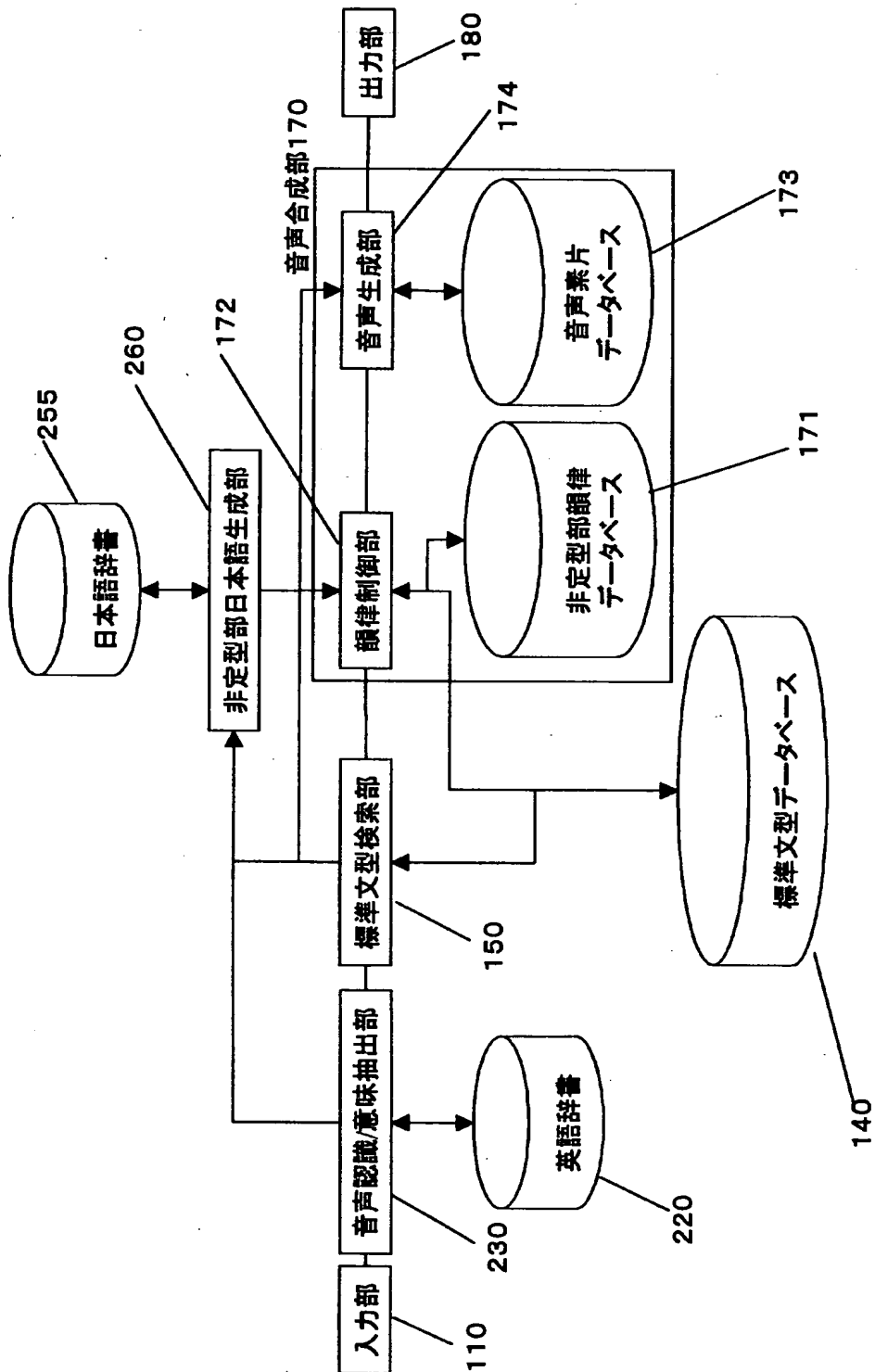
【図 2】



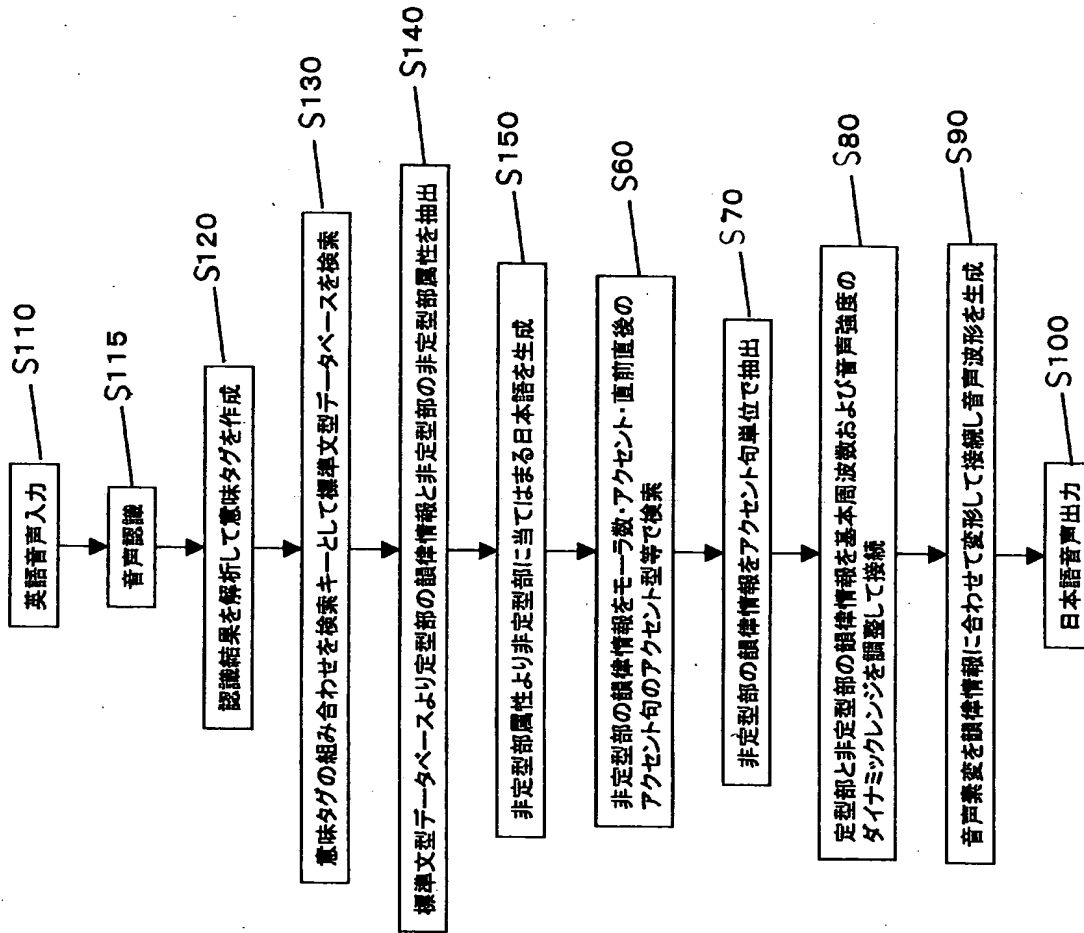
【図 3】



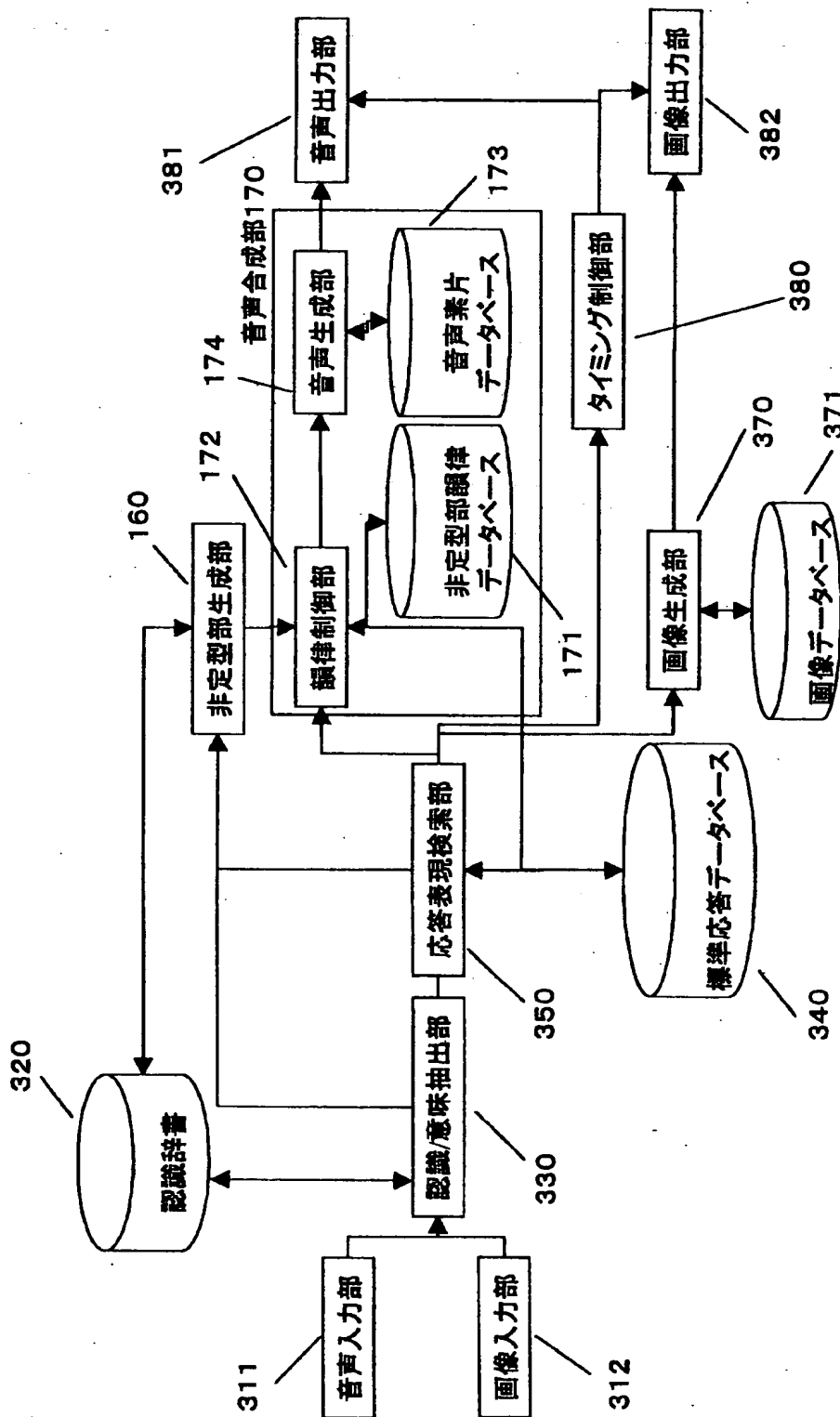
【図 4】



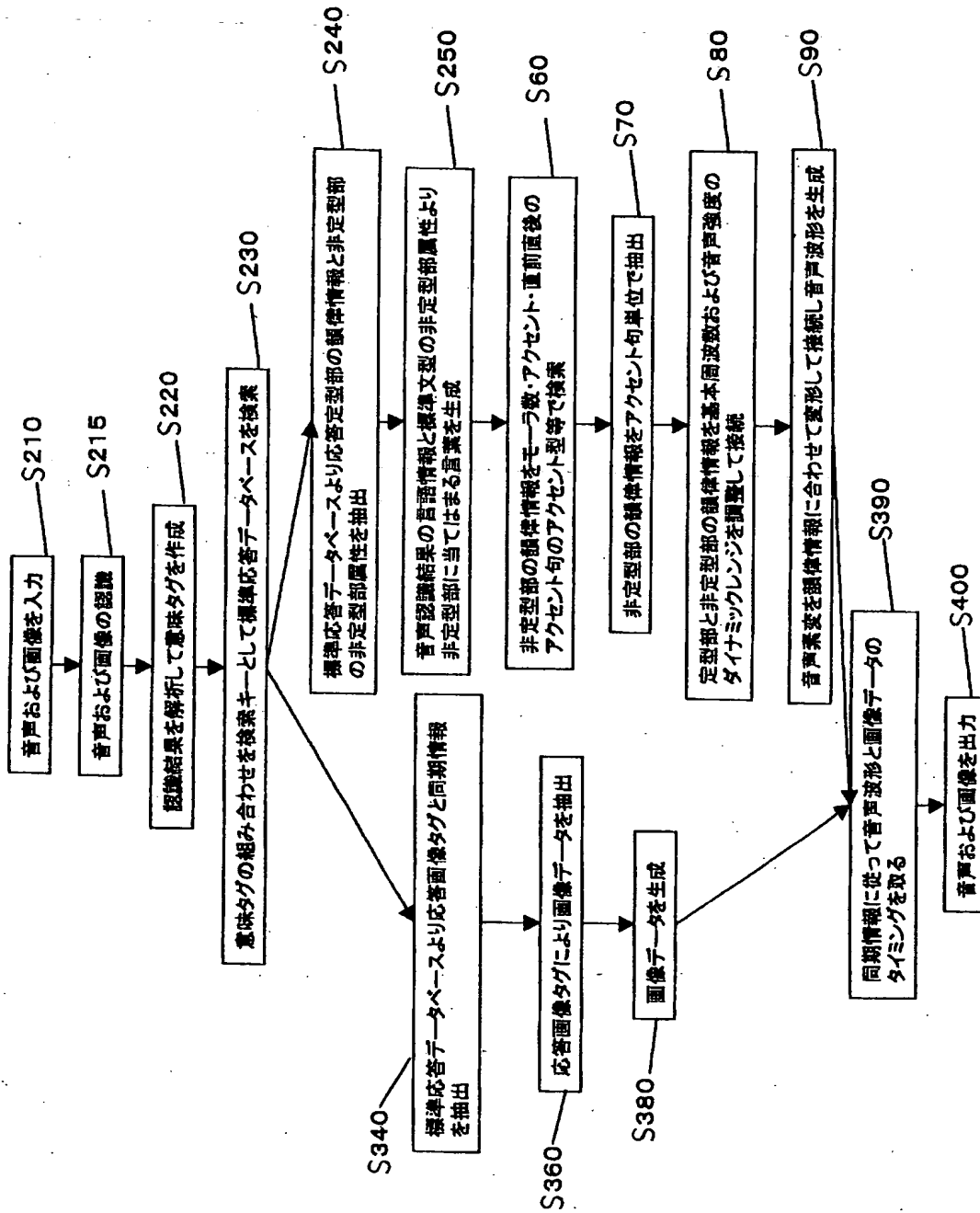
【図5】



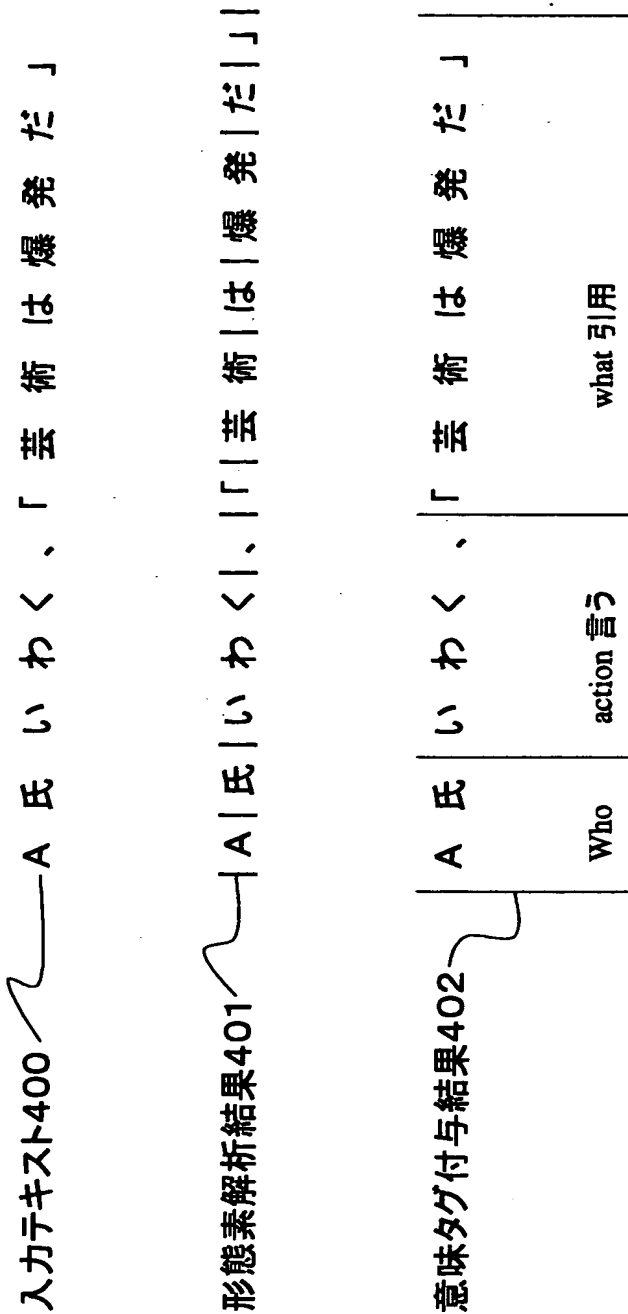
【図6】



【図 7】



【図 8】





【書類名】 要約書

【要約】

【課題】 入力信号の意味を解釈し標準文型による言語表現に変換することで広範なメディア、モダリティから音声への変換および言語変換を可能にし、かつ高品質な音声で情報を提供すること。

【解決手段】 入力信号の意味を抽出する意味抽出部 1 3 0 と、標準文型データベース 1 4 0 より入力信号から抽出された意味と同等の標準文型を検索してその標準文型の定型部に対応する音声情報を抽出する標準文型検索部 1 5 0 と、標準文型の非定型部に当てはまる言葉を入力信号の意味と非定型部の属性から生成し、非定型部韻律データベース 1 7 1 より非定型部に当てはまる言葉に対応する韻律情報を選択し、その韻律情報に従って音声情報を生成し、定型部と非定型部の音声情報を接続し、その音声情報に基づいて音声波形を生成する音声合成部 1 7 0 と、音声を出力する出力部 1 8 0 とを備える。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000005821]

1. 変更年月日	1990年 8月28日
[変更理由]	新規登録
住 所	大阪府門真市大字門真1006番地
氏 名	松下電器産業株式会社